GENERALIZED SEMI ONE-LEVEL ROTATION SAMPLING

YouSung Park, KeeWhan Kim, Korea University, and Jai Won Choi, U. S. National Center for Health Statistics
YouSung Park, Statistics Department, 5-1 Anam-Dong, Sungbuk-gu, Seoul, Korea

Key Words: Rotation sampling design, Generalized Composite estimator, Variance, Efficiency.

1. Introduction

There are two basic requirements in rotation sampling : all rotation groups are included in any given time period, and there are some sample units overlapping and other units are systematically replaced, either completely or partially, in successive time periods, according to rules of rotation. Rotation is used to avoid undue reporting burden expected from survey respondents and to obtain the information of changes from overlapping units. In semi one-level rotation design, sample units in a rotation group drops out from sample and then return to that sample later time.

In this paper, we present some formal rules for semi one-level rotation design. In this design, each primary sampling unit(psu) is divided into rotation groups in such a way rotation groups are homogeneous within psu. Each of rotation groups includes a number of clusters. These clusters are the final sampling units. The clusters within the rotation group are rotated. For instance, in U.S. Current Population Survey(CPS) there are 8 rotation groups in a psu, and each rotation group includes clusters, and each cluster consists of four households. A sample cluster in each of 8 rotation groups stays in the sample for four months, leave the sample for the next eight months. Then the same cluster returns to the sample for the following four months. We call it 4-8-4 design. From now on, we will use month for the time period.

In Section 2, we present some formal rules regulating the number of months for certain clusters to stay, leave, or return again to the sample. In order to satisfy the basic requirements, three necessary rules are proposed in this section. In Section 3, based on these formal rules in Section 2, we present algorithms and overlapping rules by which rotation groups are allocated. In Section 4, we show new designs of rotation sampling, and obtain the variances and optimal coefficients of the Generalized Composite Estimators(GCE) for each design. We also investigate the efficiency of the GCE of new design over GCE of the usual 4-8-4 design.

The authors wish to acknowledge the financial support of the Korea Research Foundation made in the program year of 1997

2. Generalized semi One-Level Rotation Design

The generalized semi one-level rotation sampling design expressed as $r_1^m - r_2^{m-1}$ is defined as follows: Some of clusters in each rotation group are interviewed for consecutive r_1 months, drop out for the next r_2 succeeding months, and return to the sample for another r_1 months. This process is repeated m times before the cluster drops out of the sample. For example $4-8-4$ design can be expressed by $4^2 - 8^1$.

We assume the followings for the generalized semi one-level rotation design :

(1) The sample size is the same for all survey month.

(2) The overlapping percentage between month t and month $t+l$, depends on only time lag l.

The CPS satisfies these assumptions. However, $4^2 - 8^1$ design uses only one cluster as a sample in each of 8 rotation groups, and estimates only year-to year change since no overlapping occurs between months t and $t+12k$ for $k \geq 2$. We now are free from such restrictions in $r_1^m - r_2^{m-1}$ design by allowing one or more clusters as a sample and overlapping to occur 2 or more years later to estimate changes. We investigate the relationship among r_1, r_2, m, and the number of rotation groups satisfying these assumptions.

When $r_1 = 0$ or $m = 0$, this implies no sample. And $r_1^m - r_2^{m-1}$ is a fixed sample design when $r_2 = 0$ or $m = 1$. Therefore, we only consider $r_1^m - r_2^{m-1}$ rotation design for $r_1 \geq 1$, $r_2 \geq 1$, and $m \geq 2$.

Theorem 1. *Suppose that for $r_1 \geq 1$, $r_2 \geq 1$, and $m \geq 2$, $r_1^m - r_2^{m-1}$ design satisfies the assumptions (1) and (2) , and has at most k years overlapping(i.e, the overlapping between months t and $t+12k$). Then* *(a) For all survey months, the number of sample clusters appearing in the sample for the first time, 2nd time,$\cdots$, and mr_1th time are all the same.*

(b) For each $k = 1, 2, \ldots$, $l = 1, 2, \ldots$, such that $r_2 = lr_1 < 12k$, m, r_1, and r_2 are determined with the constraint $12k - r_1 + 1 \leq (m-1)(r_1 + r_2) \leq 12k + 12 - r_1$ except for the cases of r_1 and r_2 satisfying $12k - r_2 - 1 < (m_0^* - 1)(r_1 + r_2) < 12k - r_1 + 1$ where $m_0^* = \{m^*; (m^*-1)(r_1+r_2) \leq 12k \leq m^*(r_1 + r_2), m^* = 1, 2, \ldots, m\}$.

(c) The necessary number of rotation groups is mr_1, unchanged in time and all rotation groups are included in the sample for any survey month if the sample clusters in a rotation group are rotated in and out simultaneously.

Proof. **(a)** By the definition of $r_1^m - r_2^{m-1}$ design, since the sample clusters can be interviewed for the maximum of mr_1 months, we can partition the sample clusters into mr_1 individual subsets by the number of appearance in the sample at any survey month. Define $\mathcal{G}_t = \{g_{t,i,j}, 1 \leq i \leq m, 1 \leq j \leq r_1\}$ where $g_{t,i,j}$ is a subset of clusters which appear in the sample $(i-1)r_1 + j$ times at a survey month t, $t = 1, 2, \ldots$. Assume that the size of $g_{t,i,j}$, the number of clusters in $g_{t,i,j}$, $n_{t,i,j}$. Consider two sets of $\mathcal{G}_t$ and $\mathcal{G}_{t+1}$. The followings can be shown easily by the definition of $r_1^m - r_2^{m-1}$ design :

(i) For each $J_1 = 1, 2, \ldots, r_1 - 1$, only $\{g_{t,i,j}\}$ and $\{g_{t+1,i,j}\}, 1 \leq i \leq m, 1 \leq j \leq r_i - J_1$ return to the sample at month $t + J_1$ and $t + 1 + J_1$, respectively. Here define $J_1 = 0$ when $r_1 = 1$.

(ii) For each $I = 1, 2, \ldots, m-1$ and $J_2 = r_1 - 1, r_1 - 2, \ldots, 0$, only $\{g_{t,i,j}\}$ and $\{g_{t+1,i,j}\}, 1 \leq i \leq I, J_2 + 1 \leq j \leq r_1$ come back to the sample at month $t + (m-I)(r_1+r_2) - J_2$ and $t + 1 + (m-I)(r_1+r_2) - J_2$, respectively.

(iii) For each $I = 1, 2, \ldots, m-1$ and $J_3 = 1, 2, \ldots, r_1 - 1$, only $\{g_{t,i,j}\}$ and $\{g_{t+1,i,j}\}, 1 \leq i \leq I, 1 \leq j \leq r_1 - J_3$ return to the sample at month $t + (m-I)(r_1+r_2) + J_3$ and $t + 1 + (m-I)(r_1+r_2) + J_3$, respectively. Define $J_3 = 0$ when $r_1 = 0$.

Therefore, the overlapping proportion between two months t and $t + J_1$, and $t+1$ and $t + J_1 + 1$ are

$$\frac{\sum_{i=1}^{m}\sum_{j=1}^{r_1 - J_1} n_{t,i,j}}{\sum_{i,j} n_{t,i,j}} \text{ and } \frac{\sum_{i=1}^{m}\sum_{j=1}^{r_1 - J_1} n_{t+1,i,j}}{\sum_{i,j} n_{t+1,i,j}} \quad (1)$$

respectively, for each J_1; the overlapping proportion between t and $t + (m-I)(r_1 + r_2) - J_2$, and $t+1$ and $t + (m-I)(r_1+r_2) - J_2 + 1$ are

$$\frac{\sum_{i=1}^{I}\sum_{j=J_2+1}^{r_1} n_{t,i,j}}{\sum_{i,j} n_{t,i,j}} \text{ and } \frac{\sum_{i=1}^{I}\sum_{j=J_2+1}^{r_1} n_{t+1,i,j}}{\sum_{i,j} n_{t+1,i,j}} \quad (2)$$

respectively, for each I and J_2; the overlapping proportion between t and $t + (m-I)(r_1+r_2) + J_3$, and $t+1$ and $t + (m-I)(r_1+r_2) + J_3 + 1$ are

$$\frac{\sum_{i=1}^{I}\sum_{j=1}^{r_1 - J_3} n_{t,i,j}}{\sum_{i,j} n_{t,i,j}} \text{ and } \frac{\sum_{i=1}^{I}\sum_{j=1}^{r_1 - J_3} n_{t+1,i,j}}{\sum_{i,j} n_{t+1,i,j}} \quad (3)$$

respectively, for each I and J_3. Since $\sum_{i,j} n_{t,i,j} = \sum_{i,j} n_{t+1,i,j}$ and the overlapping percentage depends on only time lag by the assumptions (1) and (2), recursively solving (3) from $J_3 = r_1 - 1$ to 1 for each I, we have

$$n_{t,i,j} = n_{t+1,i,j}, \ 1 \leq i \leq m-1, 1 \leq j \leq r_1 - 1. \quad (4)$$

Similarly, from (2) with together (4), we have

$$n_{t,i,r_1} = n_{t+1,i,r_1}, \ 1 \leq i \leq m-1. \quad (5)$$

Finally, the equation (1) with (4) and (5)

$$n_{t,m,j} = n_{t+1,m,j} \ 1 \leq j \leq r_1 - 1. \quad (6)$$

(4)-(6) and $\sum_{i,j} n_{t,i,j-1} = \sum_{i,j} n_{t+1,i,j}$ yield

$$n_{t,i,j} = n_{t+1,i,j} \text{ for all } i, j. \quad (7)$$

Note that since $g_{t,i,j-1} = g_{t+1,i,j}, 1 \leq i \leq m-1, 2 \leq j \leq r_1$, we have

$$n_{t,i,j-1} = n_{t+1,i,j}, \ 1 \leq i \leq m-1, 2 \leq j \leq r_1 \quad (8)$$

These (7) and (8) show that $n_{t,i,j}$'s and $n_{t+1,i,j}$'s are all the same. Since this is true for any two consecutive survey months, the claim holds.

(b) Suppose that the sample clusters in a rotation group are off for r_2 months. Then, during these r_2 months, the clusters will be replaced by new clusters and the new clusters remaining in the sample for r_1 successive months. If such replacement happens l times, $l = 1, 2, \ldots$, during these r_2 months, we have $r_2 = lr_1$ since each replaced clusters have to be surveyed for exactly r_1 successive months.

In order to get an estimate of change between month t and $t + 12k, t = 1, 2, \ldots$, we need to overlapping over k years. Whether or not the overlapping occurs depends on m, r_1, and k. From (ii) in (a), after month $t + r_1 - 1$, $\{g_{t,1,r_1}, g_{t,2,r_1}, \cdots, g_{t,m,r_1}\}$ in $\mathcal{G}_t$ appears to the sample at month $t + r_2 + 1$ for the first time. This implies that $r_2 < 12k$ since no k years overlapping occurs when $r_2 \geq 12k$. From (iii) in (a), the last appearance of $\mathcal{G}_t$ occurs at month $t + mr_1 + (m-1)r_2 - 1$. Hence, when $mr_1 + (m-1)r_2 - 1 < 12k$ we have no overlapping for k years. If $mr_1 + (m-1)r_2 - 1 \geq 12k + 12$, then we may have overlapping for $(k+1)$ years. This gives $12k - r_1 + 1 \leq$

$(m-1)(r_1+r_2) \leq 12k+12-r_1$.
The clusters only in $\{g_{t,1,1}, \cdots, g_{t,m-m_0^*+1,1}\}$ return to the sample at month $t+(m_0^*-1)(r_1+r_2)+r_1-1$, and all clusters in $\mathcal{G}_t$ do not return to the sample from the month $t+(m_0^*-1)(r_1+r_2)+r_1$ to the month $t+m_0^*(r_1+r_2)-r_1$ by (iii) in (a). And then only $g_{1r_1}, g_{2r_1}, \cdots, g_{m-m_0^*,r_1}$ appear to the sample at month $t+m_0^*(r_1+r_2)-r_1+1$. Therefore, by the definition of m_0^*, there is no overlapping between month t and $t+12k$ when $(m_0^*-1)(r_1+r_2)+r_1-1 < 12k$ and $m_0^*(r_1+r_2)-r_1+1 > 12k$.

(c) Since $n_{t,i,j}$ for $1 \leq i \leq m$, $1 \leq j \leq r_1$ and $t=1,2,\ldots$ are all the same by (a), let $n_{t,i,j} \equiv n_c$. This n_c is greater than 0 because $n_c = 0$ means no sample. Define $R_{t^*,i,j}$, $t^* = t, t+1, \ldots$ be the rotation group which contains $g_{t^*,i,j}$ as the sample clusters appearing in the sample at month t^* for the $(i-1)r_1+j$ times. Thus for the initial $t^*=t$, we have mr_1 rotation groups, $R_{t,i,j}, 1 \leq i \leq m, 1 \leq j \leq r_1$ in which each $R_{t,i,j}$ contains $g_{t,i,j}$. Since the sample clusters in a rotation group are rotated in and out simultaneously, and the sample clusters are rotated within the same rotation group, for $t^* \geq t+1$ affiliate $g_{t^*+1,1,1}$ to the rotation group R_{t^*,m,r_1} and $g_{t^*+1,i,j+1}$ to the rotation group $R_{t^*,i,j}$ for $1 \leq i \leq m-1$, $1 \leq j \leq r_1-1$. For the remaining $g_{t^*+1,i+1,1}, 1 \leq i \leq m-1$, each clusters $g_{t^*+1,i+1,1}$ are attached only one of $m-1$ rotation groups, $\{R_{t^*,i,r_1}, 1 \leq i \leq m-1\}$. This affiliation rule guarantees that the number of rotation groups is mr_1 and invariant in time because $n_c \geq 1$, and satisfies the basic requirement that all rotation groups are included in any given time period given in Section 1. Now, suppose that the number of rotation groups is less that mr_1. Then at least one rotation group contains more than one $g_{t,i,j}$ at month t. Since different $g_{t,i,j}$'s have different rotation patterns, the sample clusters in the rotation group containing more than one $g_{t,i,j}$ can not be rotated in and out simultaneously. This complete the proof. $\square$

Although the time period is month in Theorem 1, it could be a quarter or a half year. Then, simply, change $12k$ and 12 in (b) into $4k$ and 4 for a quarterly data, and $6k$ and 6 for a semiannual data.

Example 1. The following examples are only for (b). Consider 3^3-6^2 When $k=1$. Since $r_1=3, r_2=6, m=3$ and $l=2$, 3^3-6^2 satisfies $r_2 = lr_1 < 12k$ and $12k-r_1+1 \leq (m-1)(r_1+r_2) \leq 12k+12-r_1$. Observe $m_0^*=2$. It is easy that 3^3-6^2 satisfies the exception case of $12k-r_1-1 < (m_0^*-1)(r_1+r_2) < 12k-r_1+1$. Thus, 3^3-6^2 design can not be a member of our $r_1^m - r_2^{m-1}$ designs. However, one can easily check that 5^2-10^1 design satisfies (b). From (b), one can also obtain 31 rotation designs when $k=1$ such as $1^4-5^3, 2^6-2^5, 3^2-9^1, 4^2-8^1, 5^2-10^1$, and 8^2-8^1 etc. There are 51 designs for $k=2$, 57 designs for $k=3$,73 designs for $k=4$ and 82 designs for $k=5$.

3. An Algorithm how to allocate rotation groups

The algorithm in this section comes form the affiliation rule in the proof (c) of Theorem 1. More precisely the algorithm is a rule by which we determine $R_{t,i,j}$ and $R_{t^*,i,j}$, $t^* \geq t+1, 1 \leq i \leq m, 1 \leq j \leq r_1$ systematically to satisfy two assumptions (1) and (2), and the properties of Theorem 1. Since once $R_{t,i,j}$'s are determined, $R_{t^*,i,j}$'s are straightforward as discussed in the proof (a) of Theorem 1, call $\{R_{t,i,j}, 1 \leq i \leq m, 1 \leq j \leq r_1\}$ to be the first sample. To develop the algorithm, let us identify mr_1 rotation groups by numbers $1, 2, \cdots, mr_1$, and define the αth panel, P_α, to be the set of αth clusters in each of mr_1 groups. To denote that which clusters come from which rotation group, we use the rotation group numbers $1, 2, \cdots, mr_1$ in the P_α. That is, $P_\alpha = \{1, 2, \cdots, mr_1\}$.
Before developing the algorithm consisting 6 steps, take an example for easy understanding of the algorithm and the reason why we need P_α. Figure 1 illustrates 2^4-2^3 design. Since $m=4$ and $r_1=2$, we have 8 rotation groups and $P_\alpha = \{1, 2, \cdots, 8\}$. As seen in Figure 1, the first sample(Jan. of Year 1 ; $R_{t,i,j}$'s) consists of group numbers $1, 2, 5$ and 6 in the panel P_1, and $3, 4, 7$ and 8 in the panel P_2. The second sample(Feb. of Year 1 ; $R_{t+1,i,j}$'s) is determined by simply going one step to the right from the first sample. Generally, with appropriate panels, the qth sample($R_{t+q,i,j}$'s) is made by going exactly $q-1$ steps to the right from the first sample. This sampling procedure ensures the consistent overlapping and $r_1 = 2$ months on and $r_2 = 2$ months off. One can easily see that 2^4-2^3 design

Figure 1: 2^4-2^3 system

		Panels and rotation groups																													
		P_1								P_2								P_3								P_4							
YEAR	MON	1	2	3	4	5	6	7	8	3	4	5	6	7	8	1	2	1	2	3	4	5	6	7	8	3	4	5	6	7	8	1	2
Y1	JAN	o	o			o	o			o	o			o	o																		
	FEB		o	o			o	o			o	o			o	o																	
	MAR			o	o			o	o			o	o			o	o																
	APR				o	o			o	o			o	o			o	o															
	MAY					o	o			o	o			o	o			o	o														
	JUN						o	o			o	o			o	o			o	o													
	JUL							o	o			o	o			o	o			o	o												
	AUG								o	o			o	o			o	o			o	o											
	SEP									o	o			o	o			o	o			o	o										
	OCT										o	o			o	o			o	o			o	o									
	NOV											o	o			o	o			o	o			o	o								
	DEC												o	o			o	o			o	o			o	o							
Y2	JAN													o	o			o	o			o	o			o	o						
	FEB														o	o			o	o			o	o			o	o					

in Figure 1 enjoys two assumptions as well as the

properties given in Theorem 1. However, in Figure 1, when we choose $1, 2, 3, 4$ instead of $1, 2, 5, 6$ in P_1, and $5, 6, 7, 8$ instead of $3, 4, 7, 8$ in P_2 as the first sample, the 4th sample(April of Year 1) would not have rotation group number 3. This is a violation of (c) in Theorem 1. Reordering $P_2 = (34567812)$ to $P_2 = (12567834)$ also violate the (c) in Theorem 1. These observations show that the determination of $R_{t,i,j}$ and $R_{t^*,i,j}$ depends on the ordering of rotation group numbers in each panel P_α.
We specify the subscript α for the panels P_α, $\alpha = 1, \ldots, L$, where $L = l$, if $mr_1 = (m-1)r_2$ and $L = l + 1$ otherwise. $P_1, P_2, \ldots, P_L$ are the basic panels, each with mr_1 rotation groups and l comes from the equation $r_2 = lr_1$.

We now discuss how to allocate rotation groups to meet the two assumptions for $r_1^m - r_2^{m-1}$ rotation designs that also meet the conditions given in Theorem 1. This allocation procedures are explained in the following 6 steps with example for $2^4 - 2^3$ system in Figure 1.

Step 1. For P_1, allocate the rotation group numbers by increasing order. Namely, $P_1 = (1, 2, 3, \ldots, mr_1)$. In Figure 1, $P_1 = (1, 2, \ldots, 8)$.
Step 2. Write the basic panels $P_1, P_2, \ldots, P_L$, each occupying mr_1 positions. The αth panel, P_α, includes positions from the position $(\alpha - 1)mr_1 + 1$ to αmr_1. In Figure 1, $L = 2$, we have $Lmr_1 = 16$ positions, and P_2 has the positions from the 9th to the 16th.
Step 3. Indicate the symbol 'o' for the first r_1 positions. After the first successive r_2 positions with no symbol, indicate the second r_1 positions with the same symbol. And so on until mth r_1 positions are indicated. See the row of JAN.YEAR1 in Figure1.
Step 4. Fill the checked positions in $P_2, P_3, \cdots, P_L$ by turn with the group numbers that were not checked in P_1. The rotation group numbers checked in this Step 4 are the first sample.
In Figure 1, see the row of JAN.YEAR1. The checked positions, the first and second, of P_2 are filled by group numbers, 3 and 4, respectively. Other positions checked by 'o' in P_2 are the 5th and 6th positions, filled by the group numbers, 7 and 8. These 3, 4, 7 and 8 were not checked group numbers in P_1.
Step 5. Fill the remaining rotation group numbers by circular order in the empty positions of P_α, $\alpha = 2, \ldots, L$ that are partially occupied from Step 4. In Figure 1, there are 4 empty positions, the 3rd, 4th, 7th and 8th in P_2. The first 2 empty positions and the second 2 positions are filled group numbers 5 and 6, 1 and 2, respectively because the number 8 is followed 1 by circular ordering.
When the panels P_α, for $\alpha \leq L$ exist with no check mark , copy the nearest $P_{\alpha'}$, $\alpha' \leq \alpha$.
Now, we have the arrangement of L panels, P_1, $P_2, \cdots, P_L$. Copy these panels to the next L panels, $P_{L+1}, \cdots, P_{2L}$ and so on.
Step 6. With the first sample defined in Step 4 and P_α given in Step 5, $R_{t+q,i,j}$ are determined by going exactly $q - 1$ steps to the right from the first sample as shown in Figure 1.

3.1 Overlapping

Let y_t be a monthly level estimator at month t. The variance of t^* months change, $Var(y_{t+t^*} - y_t) = Var(y_{t+t^*}) - 2Cov(y_{t+t^*}, y_t) + Var(y_t)$. This means that $Var(y_{t+t^*} - y_t)$ depends on the overlapping percentage between months t and $t + t^*$ as well as sample correlation. The overlapping percentage in $r_1^m - r_2^{m-1}$ design depends on r_1, r_2, and m. From the proof (a) in Theorem 1, the overlapping percentage between months t and $t + t^*$ are as follows.

$$O(t, t^*) = \begin{cases} \frac{r_1 - t^*}{r_1} & \text{if } \; 1 \leq t^* \leq r_1 - 1 \\ \frac{(m-i)(r_1 - |j|)}{mr_1} & \text{if } \; t^* = i(r_1 + r_2) - j, \\ & \quad i = 1, \ldots, m-1, \\ & \quad j = r_1 - 1, r_1 - 2, \ldots, 1 - r_1 \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

For fixed k, the biggest monthly overlapping occurs $m = 2$. The reason is that the smaller m, the larger r_1 is by (c) of Theorem 1. From (9), the overlapping percentage between month t and $t + (r_1 + r_2)$ is $(m-1)/m$. This implies that the overlapping between month t and $t + (r_1 + r_2)$ becomes larger as m is larger. For example, when $k = 1$, the largest monthly overlapping percentage of 87.5% is obtained in $8^2 - 8^1$; the largest quarterly overlapping percentage of 87.5% in $1^8 - 2^7$; the largest 6 months overlapping percentage 75% in $1^8 - 2^7, 1^{12} - 1^{11}, 2^4 - 2^3$ and $3^4 - 3^3$; the largest yearly overlapping percentage 50% in 12 designs such as $2^4 - 2^3, 3^4 - 3^3, 3^2 - 9^1, 4^2 - 8^1$ and $7^2 - 7^1$.
For fixed sample correlation, the variance for the change between months t and $t + t^*$ decreases as t^* months overlapping increases since overlapping correlation is positive in usual sampling. Therefore, the choice of rotation design depends on amount of change we like to choose. Note that the designs with $r_1 = 1$ have no monthly overlapping and so the variance reduction for monthly change is not expected.

4. Composite Estimator

Several estimators have been developed for rotation sampling during the past several decades starting from simple composite estimators to more general forms. The latter forms are developed with fixed weighting values(Hansen,1978; Huang and Ernst, 1981; Kumer and Lee, 1983).

We quote a GCE from Cantwell (1990) and obtain its coefficients by minimizing the variance equation as follows.

Denote an estimate by $x_{t,i,j}$ for the jth cluster in the ith appearance, $j = 1,\ldots,n_c$ $i = 1,2,\cdots M$, and month $t = 1,\cdots,\infty$ where n_c is the size of clusters in the ith appearance. Since the clusters are selected by simple random sampling, define $x_{t,i} = 1/n_c \sum_{j=1}^{n_c} x_{t,i,j}$; $x_{t,i}$ measures some characteristic of the rotation group. The part of this rotation group is included in the monthly sample as the ith appearance in month t. For the system of $r_1^m - r_2^{m-1}$, we can express the number of total appearances in month t as $M = mr_1$; then we can write the GCE for month t as

$$y_t = \sum_{i=1}^{M} a_i x_{t,i} - \omega \sum_{i=1}^{M} b_i x_{t-1,i} + \omega y_{t-1}$$

where the weight ω is bounded as $0 \leq \omega < 1$. The other weights a_i's and b_i's may take any values subject to $\sum_{i=1}^{M} a_i = 1$, and $\sum_{i=1}^{M} b_i = 1$. As we expect, the GCE become to be the simple estimator when $\omega = 0$.

We assume the covariance between $x_{t,i}$ and $x_{s,j}$ $Cov(x_{t,i}, x_{s,j}) = 0$, if $i \neq j$, $Cov(x_{t,i}, x_{s,j}) = \rho_{ts}\sigma^2/n_c$, if $i = j$, and two $x_{t,i}, x_{t,j}$ are in the same rotation group, where the subscript in ρ_{ts} is $ts = |t-s|$ so that the correlation is a function only on the absolute value of time difference $t-s$, and $\rho_0 = 1$.

Under this covariance structure, following the results of (9) and Theorem 1, for $r_1^m - r_2^{m-1}$ design, it can be shown that

$$\begin{aligned} Var(y_t) =& \frac{\sigma^2}{n_c}\{\mathbf{a}'\mathbf{a} + \omega^2\mathbf{b}'(\mathbf{b} - 2\mathbf{a}) \\ &+ 2(\mathbf{a} - \omega^2\mathbf{b})'\mathbf{Q}(\mathbf{a} - \mathbf{b})\}/(1-\omega^2) \quad (10) \end{aligned}$$

where the weights $\mathbf{a}' = (\mathbf{a}_1', \mathbf{0}', \mathbf{a}_2', \mathbf{0}', \mathbf{a}_3', \cdots, \mathbf{a}_m')$, $\mathbf{b}' = (\mathbf{b}_1', \mathbf{0}', \mathbf{b}_2', \mathbf{0}', \cdots, \mathbf{b}_m')$ with the ith element $\mathbf{a}_i' = (a_{(i-1)(r_1+r_2)+1}, \cdots, a_{ir_1+(i-1)r_2})$, $\mathbf{b}_i' = (b_{(i-1)(r_1+r_2)+1}, \cdots, b_{ir_1+(i-1)r_2})$ and $\mathbf{0}$ to be a zero vector of size $r_2 \times 1$, and for any given ω, the correlation matrix $Q = (q_{i,j})_{T\times T}$, $q_{i,j} = \omega^{i-j}\rho_{i-j}$ for $1 \leq j < i \leq T$, $T = mr_1 + (m-1)r_2$. This correlation matrix Q is partially responsible for the size of this variance. Note that the variance of simple estimator is $\sigma^2/n_c\mathbf{a}'\mathbf{a}$ which is minimized at $a_i = 1/M, i = 1,\ldots,M$.

An estimator of population change from one month to the next is $d_t = y_t - y_{t-1}$. It also can be seen that for $\omega \neq 0$,

$$\begin{aligned} Var(d_t) =& \frac{\sigma^2}{n_c}(\mathbf{a}'\mathbf{a} + \omega^2\mathbf{b}'\mathbf{b} - 2\omega\rho_1\mathbf{a}'L\mathbf{b})/\omega \\ &- (1-\omega^2)Var(y_t)/\omega \quad (11) \end{aligned}$$

where $\mathbf{a}$ and $\mathbf{b}$ are given as above, and $L = (l_{ij})_{T\times T}$ with elements $l_{ij} = 1$ when $i - j = 1$, and $l_{ij} = 0$ when $i - j \neq 1$. When $k = 0$, with the variance $Var(d_t) = 2\sigma^2/n_c\mathbf{a}'(I - \rho_1 L)\mathbf{a}$. This variance is called as simple estimator of monthly change is minimized at $a_i = 1/M, i = 1,\ldots,M$.

4.1 Derivation of optimal weights a and b

Now, we find the coefficients of $\mathbf{a}$ and $\mathbf{b}$ to minimize $Var(y_t)$ and $Var(d_t)$, respectively for fixed ω .

Define

$$\begin{aligned} A_1 =& (1-\omega^2)(\alpha_c^* G - I)\left(\frac{\omega^2}{1-\omega^2}I + \alpha_c^{-1}Q\right), \\ B_1 =& (\frac{1-\omega^2}{\omega^2})(\alpha_c^* G - I)\left(\frac{\omega^2}{1-\omega^2}I + \alpha_c^{-1}Q'\right) \end{aligned}$$

where $\alpha_c = I + Q + Q'$, $\alpha_c^* = \alpha_c^{-1}G'(G\alpha_c^{-1}G')^{-1}$, and G and g are appropriate matrix and vector with size $((m-1)r_1 + 1) \times (mr_1 + (m-1)r_2)$ and $((m-1)r_2 + 1) \times 1$, respectively. Those are from the constraints of $\sum_{i=1}^{M} a_i = 1$ and $\sum_{i=1}^{M} b_i = 1$.

As usual, using lagrange multiplier technique one can show that the coefficients of $\mathbf{a}$ and $\mathbf{b}$ minimizing $Var(y_t)$ are

$$\widehat{\mathbf{a}} = [I - A_1B_1]^{-1}[I - A_1]\alpha_c^* g \quad (12)$$

$$\widehat{\mathbf{b}} = \alpha_c^* g - B_1\widehat{\mathbf{a}}. \quad (13)$$

Similarly, the optimal coefficients, $\mathbf{a}$ and $\mathbf{b}$ for $Var(d_t)$ are

$$\begin{aligned} A_2 =& \alpha_m^{-1}(I - G'\alpha_m^* G\alpha_m^{-1})(\alpha_m + \beta_m + \gamma_m), \\ B_2 =& \frac{1}{\omega^2}\alpha_m^{-1}(I - G'\alpha_m^* G\alpha_m^{-1})(\alpha_m + \beta_m + \gamma_m)' \end{aligned}$$

where $\alpha_m = \frac{2\omega}{1+\omega}I - \frac{1-\omega}{1+\omega}(Q+Q')$, $\alpha_m^* = [G\alpha_m^{-1}G']^{-1}$ $\beta_m = (1-\omega)^2 Q' - \omega(2-\omega)I$, $\gamma_m = \omega\rho L$.

$$\widehat{\mathbf{a}^*} = [I - A_2B_2]^{-1}[I + B_2]\alpha_m^{-1}G'\alpha_m^* g \quad (14)$$

$$\widehat{\mathbf{b}^*} = B_2\widehat{\mathbf{a}^*} + \alpha_m^{-1}G'\alpha_m^* g. \quad (15)$$

4.2 Comparison of semi one-level rotation designs with one year overlapping

Since the simple estimator is a special case of the GCE as shown in section 4, one can expect that the variance of the GCE is smaller than that of simple estimator. For the numerical example, we assume the covariance structure given in Section 4 to have an exponential pattern as

$$\rho_{ts}\frac{\sigma^2}{n_c} = \rho^{|t-s|}\frac{\sigma^2}{n_c}, \quad s = 1, 2, \ldots, \infty.$$

Since the variances (10) and (11) with optimal coefficients given (12)-(15) are a function of only ω, ω minimizing (10) and (11) was chosen from 0.1 to 0.9 with increment 0.1 in this section.

The numbers in Table 1 indicates the ratio of the variance of GCE to that of simple estimator for some selected rotation designs when $r_1 = 1$. The comparison in Table 1 is conducted for monthly change. As we expect in section 3.1, the GCE for monthly change achieves negligible amounts of variance reduction when it compares to the variance of simple estimator. This is mainly because of no monthly overlapping when $r_1 = 1$.

Table 1: Comparison for variance of GCE (when $r_1 = 1$ and $k = 1$) to the variance of simple estimator(optimal ω).

Designs	ρ 0.5	0.6	0.7	0.8	0.9
$1^2 - 11^1$	1.000(0.8)	1.000(0.8)	1.000(0.8)	1.000(0.8)	1.000(0.8)
$1^7 - 1^6$	0.998(0.1)	0.996(0.1)	0.992(0.1)	0.987(0.1)	0.977(0.1)
$1^8 - 1^7$	0.998(0.1)	0.996(0.1)	0.993(0.1)	0.988(0.1)	0.980(0.1)
$1^8 - 2^7$	1.000(0.4)	1.000(0.4)	1.000(0.4)	1.000(0.4)	0.999(0.4)
$1^{12} - 1^{11}$	0.999(0.1)	0.998(0.1)	0.995(0.1)	0.992(0.1)	0.986(0.1)

We now define the efficiency of alternative rotation designs, new design *.vs.* $4^2 - 8^1$ design when $r_1 \geq 2$. This efficiency is also defined by the rate of two variances, one from the new design the other from $4^2 - 8^1$ design. Since we have two estimators, y_t and d_t, from each design, we compare the sum of these two variances as

$$\begin{array}{c}\text{Efficiency of}\\ \text{alternative design}\end{array} = \frac{Var(y_{at}) + Var(d_{at})}{Var(y_{ut}) + Var(d_{ut})} \quad (16)$$

where y_{at} and d_{at} are GCE of alternative rotation design for monthly level and month to month change, respectively, and y_{ut} and d_{ut} are composite estimators for $4^2 - 8^1$ design. Although we did not present the separate results of $Var(y_{at})/Var(y_{ut})$ and $Var(d_{at})/Var(d_{ut})$, it is worth to observe the following. For monthly level with fixed ρ, the rotation designs with smaller monthly overlapping have smaller variance ; on the other hand, for month to month change with fixed ρ, the rotation designs with larger monthly overlapping have smaller variance.
The numbers in Table 2 are the efficiency defined in (16) with the same sample size 64. Smaller value less than 1 of efficiency implies that the design in the first column is better than $4^2 - 8^1$ design for given correlation. From $2^2 - 10^1$ to $3^4 - 3^3$, all of them are worse than $4^2 - 8^1$ except for $3^2 - 9^1$ when $\rho = 0.9$. All designs with $r_1 > 4$, i.e, from $5^2 - 5^1$ to $8^2 - 8^1$, all of them are better than $4^2 - 8^1$ for $\rho = 0.5, 0.6, 0.7$. There is no design consistently better than $4^2 - 8^1$ for all ρ. However, one may choose last 5 designs, especially, $5^2 - 10^1$ as we may ignore, $\rho = 0.9$ is an exceptional case in practice. The main reason is that, as shown in (9), the monthly overlapping increase as r_1 is bigger. The results of Table 2 may be changed when we consider more than monthly overlapping. If we assume seasonal correlation pattern, we scrutinize lots of interesting properties of our generalized rotation designs. However all of these require derivations of variance for GCE of more than one month change and optimal coefficients to minimize the variances. These study will be in later paper.

Table 2: Efficiency of alternative designs under exponential correlation pattern

Designs	ρ 0.5	0.6	0.7	0.8	0.9
$2^2 - 10^1$	1.0787	1.0900	1.0951	1.0881	1.0488
$2^3 - 4^2$	1.0786	1.0897	1.0951	1.0913	1.0697
$2^4 - 2^3$	1.0779	1.0895	1.0981	1.1046	1.1195
$2^4 - 4^3$	1.0786	1.0897	1.0951	1.0917	1.0725
$2^5 - 2^4$	1.0778	1.0894	1.0981	1.1049	1.1228
$2^6 - 2^5$	1.0777	1.0893	1.0981	1.1051	1.1245
$3^2 - 9^1$	1.0212	1.0223	1.0207	1.0134	0.9916
$3^3 - 3^2$	1.0211	1.0226	1.0237	1.0253	1.0435
$3^4 - 3^3$	1.0211	1.0227	1.0238	1.0264	1.0492
$5^2 - 5^1$	0.9895	0.9900	0.9931	1.0024	1.0325
$5^2 - 10^1$	0.9895	0.9900	0.9923	0.9985	1.0127
$6^2 - 6^1$	0.9834	0.9848	0.9899	1.0037	1.0441
$7^2 - 7^1$	0.9795	0.9818	0.9891	1.0066	1.0537
$8^2 - 8^1$	0.9768	0.9800	0.9893	1.0105	1.0642

REFERENCES

Breau,P., and Ernst,L.(1983). Alternative estimators to the current composite estimators. *Proceedings of the Section on Survey Research Methods, American Statistical Association,* 397-402.

Huang,E.T., and Ernst,L.R.(1981). Comparison of an alternative estimator to the current composite estimator in in CPS.*Proceedings of the Section on Survey Research Methods, American Statistical Association,* 303-308.

Kumar,S., and Lee,H.(1983). Evaluation of composite estimation for the Canadian Labor Force Survey. *Proceedings of the Section on Survey Research Methods, American Statistical Association,* 403-408.

Patrick J. Cantwell.(1990). Variance formulae for composite estimators in rotation designs. *Survey Methodology.* 16, 153-163.